# CST1204: Introduction to Databases

Jun Shan
Adjunct Professor
City Tech, CUNY

# Week 10 Session 1

# 10/28/2019

# Homework Review

- Chapter 4 exercise questions for TAL Distributors 16-18
- Chapter 8 exercise questions for Colonial 1-3

# Agenda

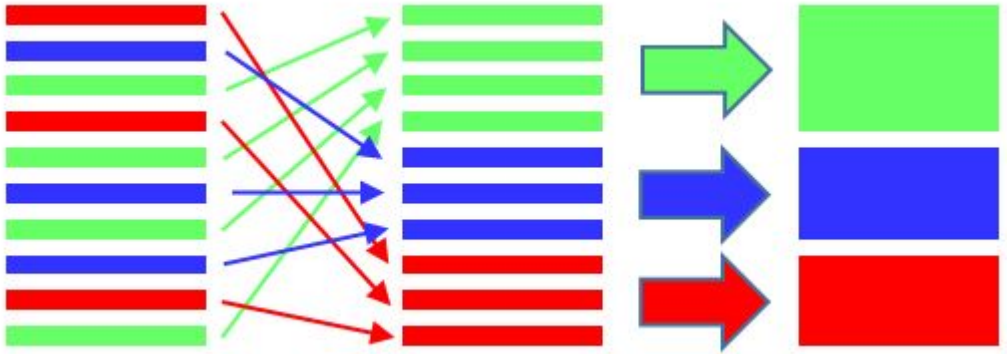
- GROUP BY

# Purpose of GROUP BY

- Aggregation functions: SUM, AVG, MIN, MAX, COUNT, COUNT DISTINCT
  - Describing characteristics of a dataset
- The need to divide a dataset into smaller groups for research
  - Average grade in one course vs average grade in another course
  - Weather data in east coast cities vs data in west coast cities
  - Stock price patterns in one sector vs patterns in another sector

# GROUP BY

- Group rows in a dataset into subsets (Ch4, Pg120)
  - Putting rows of same characteristics into same processing unit



# GROUP BY without Aggregation

- SELECT … GROUP BY without functions
  - Matching SELECT with GROUP BY: If selected, columns that are not in GROUP BY are returned as repeating groups, which won't be 1NF compliant. So, SQL does not allowing selecting columns directly without GROUP BY first.

    Example: SELECT Customer_Num when GROUP BY Rep_Num

# GROUP BY with Aggregation

- SELECT … GROUP BY with functions

  When not in GROUP BY, columns are SELECTed as repeating groups. SQL allows aggregations (SUM, AVG, MIN, MAX, COUNT) on these repeating groups because these functions will only return one value. Thus it is 1NF compliant.

  SELECT Rep_Num, COUNT(Customer_Num) when GROUP BY Rep_Num

# GROUP BY

- Multi-column GROUP BY
- GROUP BY work with WHERE
- GROUP BY with ORDER BY

# HAVING

- HAVING Clause: Conditions applied to the aggregated dataset
  - HAVING vs WHERE:
    - WHERE is applied on original dataset, thus cannot operate on the aggregated dataset
    - Having is applied on the aggregate dataset
    - Use WHERE whenever possible for performance reason

# GROUP BY

```
SELECT <group by columns>, <aggregated columns>
FROM <table>
WHERE <boolean expression of table columns>
GROUP BY <group by columns>
HAVING <boolean expression of group by columns>
ORDER BY <subset of group by columns>
```

# Hands On

- Chapter 4 exercise questions for TAL Distributors 19-23

# Homework

- Chapter 4 exercise questions for TAL Distributors 19-23